Loop corrections for message passing algorithms in continuous variable models

Bastian Wemmenhove, Bert Kappen
Department of Biophysics,
SNN, Radboud University
Nijmegen, 6525 EZ Nijmegen, the Netherlands

February 1, 2008

Abstract

In this paper we derive the equations for Loop Corrected Belief Propagation on a continuous variable Gaussian model. Using the exactness of the averages for belief propagation for Gaussian models, a different way of obtaining the covariances is found, based on Belief Propagation on cavity graphs. We discuss the relation of this loop correction algorithm to Expectation Propagation algorithms for the case in which the model is no longer Gaussian, but slightly perturbed by nonlinear terms.

1 Introduction

Message passing techniques in graphical models allow for the computation of (approximate) marginal probabilities in a time interval scaling polynomially in the model size. Their discovery has consequently revolutionized several fields of applications in the past years, of which error correcting codes and vision are probably the most prominent examples. In many cases, the corresponding graphs are loopy, implying either that the error resulting from the application of loopy belief propagation (BP) is negligible for the particular model, or it can be tolerated for the particular purpose BP serves. In other cases more sophisticated refinements of BP are necessary, taking into account (part of) the loop errors.

Finding the optimal treatment of these “loop errors” motivates an active field of research, in which different solutions applying to different model classes are developed. For models involving many short loops, like on regular lattices, CVM type approaches work well [2], or tree EP approaches [3]. The latter may also be applied to correct for an incidental large loop. Unifying frameworks like the Region graphs of [4] lead to general strategies for selecting the basic clusters underlying such approaches for general model classes.

A recent analysis has shown that the local update equations of BP may be interpreted as the zero order term of an expansion in “cavity connected correlations”. These quantities are parameterizations of the “cavity distributions”, i.e., the distribution over neighbor variables of a central variable which has been removed from the graph. The Bethe approximation and BP are recovered when this cavity distribution is assumed to factorize, whereas the first order correction to the local update equations is obtained when one takes into account the pair cumulants [5]. Estimation of these pair cumulants is possible with extra runs of BP, allowing for new polynomial time algorithms, reducing errors to order $1/N^{k+1}$ when applying algorithms of which running time scales with an extra factor of N^k [6]. Although this scaling seems heavy, the large benefit of the approach is that it does not require selection of basic clusters or underlying tree-structures, since it takes into account the effect of all loops that contribute to nontrivial correlations in the cavity distribution at once.

The above “loop correction” strategy is applicable in the class of models where a perturbative expansion around the Bethe approximation makes sense, i.e., in models with large loops and relatively weak interactions. The principal requirement is that the magnitude of pair variable cumulants of cavity distributions is an order smaller than the single variable cumulants, and third order cumulants are even smaller, etc. However, heuristics based on the strategy allow for other good algorithms performing well outside these parameter regimes [7].

So far the approach has been developed for discrete variable models on a more abstract [5, 6] versus practical level [7]. In this paper we apply the idea to graphical models for continuous variables. We derive the loop corrected belief propagation equations for simple tractable Gaussian models, yielding a message passing scheme that, besides the correct average marginals, also yields the correct variances. Besides that we discuss some approaches potentially applicable to cases in which extra function approximations are necessary, and the relation with expectation propagation. A by-product of our loop corrected belief propagation equations is an algorithm that calculates exact covariance matrices for Gaussian models like the one discussed in [1], but without explicitly using linear response.

2 General ideas

The error in the result of message passing techniques that are based on local approximations for variables that interact on a graph, like belief propagation, may be viewed from two perspectives:

- The error of the Bethe-approximation is due to the fact that loops in the graph are neglected, such that nontrivial correlations between two neighbors of one variable are neglected.

- The error is due to the fact that the functional parameterization of the local marginals is not rich enough, such that it can at most be an approximation.

These two viewpoints may be argued to have the same meaning in the end, but nevertheless may lead to different strategies in the optimization of the approximation, or the improvement of the results. If the second viewpoint is the starting point for algorithms like expectation propagation, the first may be seen as the basic view for loop correction strategies.

Since expectation propagation applies well to continuous variable cases, but loop correction schemes in the sense of [5, 7, 6] have not been applied to continuous variable cases, it might be instructive to derive corresponding equations and compare them to expectation propagation approaches.

With this motivation in mind, we will firstly analyze a loop correction scheme to BP in Gaussian models. Given this scheme, we will discuss possible generalizations suitable for cases in which the model is no longer tractable.

2.1 Model introduction

The model which we will initially consider is a Gaussian model of N interacting variables, denoted by $\boldsymbol{\sigma} \in \mathrm{I\!R}^{\mathrm{N}}$ of which the total probability distribution is given by

$$\begin{aligned} P(\boldsymbol{\sigma}) &= Z^{-1} \prod_{i=1}^{N} \psi_i(\sigma_i) \prod_{j<k}^{N} \psi_{jk}(\sigma_j, \sigma_k) \\ \psi_i(\sigma_i) &= \exp\left[-\frac{1}{2s_i}(\sigma_i - \mu_i)^2\right] \\ \psi_{jk}(\sigma_j, \sigma_k) &= \exp\left[J_{jk}\sigma_j\sigma_k\right] \end{aligned} \tag{1}$$

thus the variables i have their own Gaussian local potential with average μ_i and variance s_i, but interact in a pairwise manner with variables j via the interaction J_{ij}. Obviously, $Z = \int d\boldsymbol{\sigma} \prod_{i=1}^{N} \psi_i(\sigma_i) \prod_{j<k}^{N} \psi_{jk}(\sigma_j, \sigma_k)$. We will denote the neighborhood of variable i on the graph by ∂_i, i.e. $\partial_i = \{j | J_{ij} \neq 0\}$.

2.2 The "cavity equations"

The following analysis will be based on the loop correction equations of [5], which were applied to discrete binary variables. The current generalization to continuous variables is a straightforward application of these ideas. We write down an expression for the joint probability $P^{(ij)}(\sigma_i, \sigma_j)$ of variables σ_i and σ_j on the model from which the interaction J_{ij} has been removed in two different ways. The first is in terms of the *cavity distribution* of variable i, $P^{(i)}(\sigma_{\partial_i})$, i.e., the joint distribution over the neighbors of i in the model from which i has been removed, and the second in terms of the cavity distribution of variable j, $P^{(j)}(\sigma_{\partial_j})$:

$$P^{(ij)}(\sigma_i, \sigma_j) = \frac{1}{Z_1} \int \mathrm{d}\boldsymbol{\sigma}_{\partial_i \backslash j} P^{(i)}(\boldsymbol{\sigma}_{\partial_i}) \exp\left[-(\sigma_i - \mu_i)^2/(2s_i) + \sigma_i \sum_{l \in \partial i \backslash j} J_{il}\sigma_l\right] \tag{2}$$

$$P^{(ij)}(\sigma_i,\sigma_j) \quad = \quad \frac{1}{Z_2}\int \mathrm{d}\boldsymbol{\sigma}_{\partial_j\setminus i}P^{(j)}(\boldsymbol{\sigma}_{\partial_j})\exp\left[-(\sigma_j-\mu_j)^2/(2s_j)+\sigma_j\sum_{l\in\partial_j\setminus i}J_{jl}\sigma_l\right] \qquad (3)$$

With respect to this marginal distribution two ways of writing the moment

$$\langle\sigma_i\rangle^{(ij)} \equiv \int \mathrm{d}\sigma_i\sigma_i P^{(ij)}(\sigma_i,\sigma_j) \qquad (4)$$

are

$$\begin{aligned}
\langle\sigma_i\rangle^{(ij)} &= \frac{1}{Z_1}\int \mathrm{d}\boldsymbol{\sigma}_{\partial i}\mathrm{d}\sigma_i P^{(i)}(\boldsymbol{\sigma}_{\partial i})\sigma_i\exp\left[-(\sigma_i-\mu_i)^2/(2s_i)+\sigma_i\sum_{l\in\partial i\setminus j}J_{il}\sigma_l\right]\\
\langle\sigma_i\rangle^{(ij)} &= \frac{1}{Z_2}\int \mathrm{d}\boldsymbol{\sigma}_{\partial j}\mathrm{d}\sigma_j P^{(j)}(\boldsymbol{\sigma}_{\partial j})\sigma_i\exp\left[-(\sigma_j-\mu_j)^2/(2s_j)+\sigma_j\sum_{l\in\partial j\setminus i}J_{jl}\sigma_l\right] \qquad (5)
\end{aligned}$$

which may be written in terms of effective measures

$$\begin{aligned}
\langle f(\boldsymbol{\sigma}_{\partial_i})\rangle_{i\to j} &\equiv Z_{i\to j}^{-1}\int \mathrm{d}\boldsymbol{\sigma}_{\partial i}f(\boldsymbol{\sigma}_{\partial_i})P^{(i)}(\boldsymbol{\sigma}_{\partial i})\exp\left[\mu_i\sum_{l\in\partial i\setminus j}J_{il}\sigma_l+\frac{s_i}{2}\sum_{l,k\in\partial i\setminus j}J_{il}J_{ik}\sigma_l\sigma_k\right]\\
\langle f(\boldsymbol{\sigma}_{\partial j})\rangle_{j\to i} &\equiv Z_{j\to i}^{-1}\int \mathrm{d}\boldsymbol{\sigma}_{\partial j}f(\boldsymbol{\sigma}_{\partial j})P^{(j)}(\boldsymbol{\sigma}_{\partial j})\exp\left[\mu_j\sum_{l\in\partial j\setminus i}J_{jl}\sigma_l+\frac{s_j}{2}\sum_{l,k\in\partial j\setminus i}J_{jl}J_{jk}\sigma_l\sigma_k\right]
\end{aligned} \qquad (6)$$

where $Z_{i\to j}$ and $Z_{j\to i}$ are the corresponding normalization constants. In terms of these measures, the equations (5) lead to

$$\langle\sigma_i\rangle_{j\to i} = \mu_i + s_i\sum_{l\in\partial i\setminus j}J_{il}\langle\sigma_l\rangle_{i\to j} \qquad (7)$$

The above procedure may be repeated for all other moments of the distribution $P^{(ij)}(\sigma_i)$, e.g.

$$\langle\sigma_i^2\rangle_{j\to i} \quad = \quad s_i+\mu_i^2+2s_i\mu_i\sum_{l\in\partial i\setminus j}J_{il}\langle\sigma_l\rangle_{i\to j}+s_i^2\sum_{l,k\in\partial i\setminus j}J_{il}J_{ik}\langle\sigma_l\sigma_k\rangle_{i\to j} \qquad (8)$$

etc. The moments of the true marginal distributions are integrals with respect to different measures, e.g.:

$$\langle\sigma_i\rangle = \mu_i + s_i\sum_{l\in\partial i}J_{il}\langle\sigma_l\rangle_i \qquad (9)$$

with $\langle f(\boldsymbol{\sigma})\rangle = \int d\boldsymbol{\sigma} P(\boldsymbol{\sigma}) f(\boldsymbol{\sigma})$ and

$$\langle f(\boldsymbol{\sigma}_{\partial_i})\rangle_i \equiv Z_i^{-1} \int \mathrm{d}\boldsymbol{\sigma}_{\partial i} f(\boldsymbol{\sigma}_{\partial_i}) P^{(i)}(\boldsymbol{\sigma}_{\partial i}) \exp\left[\mu_i \sum_{l\in\partial i} J_{il}\sigma_l + \frac{s_i}{2} \sum_{l,k\in\partial i} J_{il}J_{ik}\sigma_l\sigma_k\right] \quad (10)$$

All these measures reduce to functions of the above mentioned cavity distributions, which are the unknown functions of interest. It is clear however, that so far we have not specified enough local equations to solve for the full cavity distributions $P^{(i)}(\boldsymbol{\sigma}_{\partial_i})$. If we restrict ourselves, for the moment, to Gaussian models, we will be able to perform the integrations and find exact local message passing equations. We note that for a more general type of model, such local computations will be insufficient, but may be used as a basis for an approximation when an appropriate set of approximating functions is chosen, characterized by a finite set of parameters.

3 Gaussian cavity distributions

Notice that the belief propagation is recovered when one chooses to approximate the cavity distribution by a factorizing one, i.e. $P^{(i)}(\boldsymbol{\sigma}_{\partial_i}) \sim \prod_{j\in\partial_i} Q^{(i)}(\sigma_j)$. This parameterization includes the exact result when the graph is a tree, since then there can be no nontrivial correlations between variables in any cavity set ∂_i when i is absent. When there are loops in the graph, corrections to this parameterization are desirable. Various parameterizations of these corrections are possible in principle, and in [5] it was suggested to expand the cavity distributions in the cumulants, an expansion that is appropriate when either interactions are weak or loops are long.

For a Gaussian model, the cavity distributions are completely specified by their averages and covariances, such that including the second order cumulants (the first order correction to belief propagation) yields exact equations. In the following we investigate the structure of the corresponding equations and identify the exact correction to Gaussian belief propagation. An appropriate (and exact) parameterization of the cavity distribution is

$$P^{(i)}(\boldsymbol{\sigma}_{\partial i}) \sim \exp\left[-\frac{1}{2}(\boldsymbol{\sigma}_{\partial_i} - \mathbf{m}^i)^T [D_i + A_i]^{-1} (\boldsymbol{\sigma}_{\partial_i} - \mathbf{m}^i)\right] \quad (11)$$

where we have decomposed the covariance matrix in a diagonal part (D_i) and an off-diagonal part (A_i), both having the dimensions of the cavity set. The Bethe approximation, for which cavity distributions factorize, corresponds to neglecting the off-diagonal components A_i. The matrices D_i and vectors $\mathbf{m}^i$ are found through consistency equations. In the following we will denote the vector $\mathbf{J}_i$ (again the dimensions of vectors $\mathbf{m}^i$ and $\mathbf{J}_i$ are equal to that of the cavity set, $|\partial_i|$) for which $J_{ij} = 0$ as $\mathbf{J}_i^j$. The consistency equations (7), by Gaussian integration, are

found to be

$$\begin{aligned}
&\left\{[(D_j + A_j)^{-1} - s_j \mathbf{J}_j^i \mathbf{J}_j^{iT}]^{-1}[(D_j + A_j)^{-1}\mathbf{m}^j + \mu_j \mathbf{J}_j^i]\right\}_i \\
&= \mu_i + s_i \mathbf{J}_i^{jT} \left\{[(D_i + A_i)^{-1} - s_i \mathbf{J}_i^j \mathbf{J}_i^{jT}]^{-1}[(D_i + A_i)^{-1}\mathbf{m}^i + \mu_i \mathbf{J}_i^j]\right\}
\end{aligned} \tag{12}$$

From the relations of the variances, equation (8), we find:

$$\{[(D_j + A_j)^{-1} - s_j \mathbf{J}_j^i \mathbf{J}_j^{iT}]^{-1}\}_{ii} = s_i + s_i^2 \mathbf{J}_i^{jT}[(D_i + A_i)^{-1} - s_i \mathbf{J}_i^j \mathbf{J}_i^{jT}]^{-1}\mathbf{J}_i^j \tag{13}$$

For each cavity distribution D_j and $\mathbf{m}^j$ the number of pairs of equations is equal to the number of variables in the cavity set. Thus, given a covariance matrix A, the diagonals D can be determined with the second equation, and subsequently the average values $\mathbf{m}$ can be determined with the first equation. The marginal distributions then follow directly, since all variables are now known. Substituting (11) into (9), we find

$$\langle \sigma_i \rangle = \mu_i + s_i \mathbf{J}_i \left[(D_i + A_i)^{-1} - s_i \mathbf{J}_i \mathbf{J}_i^T\right]^{-1} \left[(D_i + A_i)^{-1}\mathbf{m}^i + \mu_i \mathbf{J}_i\right] \tag{14}$$

and for the second moment

$$\langle \sigma_i^2 \rangle = \langle \sigma_i \rangle^2 + s_i \left\{1 + s_i \mathbf{J}_i^T [(D_i + A_i)^{-1} - s_i \mathbf{J}_i \mathbf{J}_i^T]^{-1} \mathbf{J}_i\right\} \tag{15}$$

The only obstacle in solving these *exact* equations is yet obtaining the off-diagonal covariances A_i for each cavity set ∂_i.

Simply neglecting them, setting $A_i = 0$, we should recover the BP equations for the Gaussian model.

Using response propagation it is possible to estimate the covariances, which leads to an improvement in the results when they are small for the binary case [5, 6]. In the Gaussian case, where results from response propagation are exact, [1, 8], this procedure should thus yield exact results provided response propagation and belief propagation both converge.

4 Loop corrected belief propagation

Using the identity

$$[A + XBX^T]^{-1} = A^{-1} - A^{-1}X(B^{-1} + X^T A^{-1} X)^{-1} X^T A^{-1} \tag{16}$$

we may write

$$[(D_i + A_i)^{-1} - s_i \mathbf{J}_i^j \mathbf{J}_i^{jT}]^{-1} = D_i + A_i + \frac{(D_i + A_i)\mathbf{J}_i^j \mathbf{J}_i^{jT}(D_i + A_i)}{1/s_i - \mathbf{J}_i^{jT}(D_i + A_i)\mathbf{J}_i^j} \tag{17}$$

Defining

$$\begin{aligned}
\alpha_i^j &\equiv \mathbf{J}_i^{jT}(D_i + A_i)\mathbf{J}_i^j && (18)\\
\alpha_i &\equiv \mathbf{J}_i^T(D_i + A_i)\mathbf{J}_i && (19)\\
\epsilon_j^i &\equiv [(D_j + A_j)\mathbf{J}_j^i]_i = [A_j\mathbf{J}_j^i]_i && (20)
\end{aligned}$$

and writing v_k^i for the (diagonal) entries of D_i where k runs over ∂_i, we find that equation (13) yields

$$v_i^j + \frac{s_j}{1 - s_j\alpha_j^i}(\epsilon_j^i)^2 = \frac{s_i}{1 - s_i\alpha_i^j} \tag{21}$$

After similar simplification of equation (12), the updates for the message variances and averages become

$$\begin{aligned}
v_i^j &= \frac{s_i}{1 - s_i\alpha_i^j} - \frac{s_j}{1 - s_j\alpha_j^i}(\epsilon_j^i)^2 && (22)\\
m_i^j &= \frac{s_i}{1 - s_i\alpha_i^j}\left[\frac{\mu_i}{s_i} + \sum_{l\in\partial i\backslash j} J_{il}m_l^i\right] - \frac{s_j\epsilon_j^i}{1 - s_j\alpha_j^i}\left[\frac{\mu_j}{s_j} + \sum_{l\in\partial j\backslash i} J_{jl}m_l^j\right] && (23)
\end{aligned}$$

and the final marginals are given by

$$\begin{aligned}
v_i &= \frac{s_i}{1 - s_i\alpha_i} && (24)\\
m_i &= v_i\left[\frac{\mu_i}{s_i} + \sum_{l\in\partial i} J_{il}m_l^i\right] && (25)
\end{aligned}$$

Indeed the BP equations follow for $A_i = 0$, since in that case $\epsilon_i^j = 0$, $\alpha_i^j = \sum_{k\in\partial_i\backslash j} J_{ik}^2 v_k^i$ and $\alpha_i = \sum_{j\in\partial_i} J_{ij}^2 v_j^i$, such that the equations (modulo a transformation) reduce to the ones in [8].

The above equations allow one to explicitly interpret the meaning of the belief propagation messages, and write down expressions for their error. Indeed the messages in equation (23) represent averages and variances of cavity distributions, i.e., of the model in absence of a variable. An interesting side result in this respect comes from the observation in [8] that the averages calculated via belief propagation are exact when the algorithm converges. It follows that the message m_i^j calculated via equation (23) is equal to m_i on a graph from which variable j is removed calculated via ordinary belief propagation. In the next section we use this observation, together with similar arguments, to obtain some more exact results from ordinary belief propagation variables alone.

5 An alternative way to calculate the error in v_i for Gaussian models

The form of the loop corrected belief propagation equations imposes a relation between the BP errors in v_i and the messages m_l^i for Gaussian models. Comparing the result of equations (25) with and without cavity covariances, one may show that

$$m_i^{\mathrm{LC}} = m_i^{\mathrm{BP}} + v_i^{\mathrm{LC}} \left(m_i^{\mathrm{BP}} [\alpha_i^{\mathrm{LC}} - \alpha_i^{\mathrm{BP}}] + \sum_{l \in \partial_i} J_{il} [m_l^{i\ \mathrm{LC}} - m_l^{i\ \mathrm{BP}}] \right) \tag{26}$$

Now, since the BP averages are exact whenever BP converges [8],

$$m_i^{\mathrm{BP}} [\alpha_i^{\mathrm{LC}} - \alpha_i^{\mathrm{BP}}] = - \sum_{l \in \partial i} J_{il} [m_l^{i\ \mathrm{LC}} - m_l^{i\ \mathrm{BP}}] \tag{27}$$

Due to the interpretation of the message averages as cavity parameters, we have access to $m_l^{i\ \mathrm{LC}}$, since

$$m_l^{i\ \mathrm{LC}} = m_l^{(i)\ \mathrm{BP}} \tag{28}$$

i.e. this is the average of variable l on the graph without i, which may be obtained by running BP on the graph without variable i. Thus by running BP on the original graph once and running it on the graph without i, we can calculate v_i^{LC} by using equation (25) and writing

$$v_i^{\mathrm{LC}} = \frac{s_i}{1 - s_i \left[\alpha_i^{\mathrm{BP}} + \left(m_i^{\mathrm{BP}}\right)^{-1} \sum_{l \in \partial i} J_{il} [m_l^{(i)\ \mathrm{BP}} - m_l^{i\ \mathrm{BP}}] \right]} \tag{29}$$

provided that $m_i^{\mathrm{BP}} \neq 0$. Similar considerations (see appendix) lead to a procedure for calculating the entire covariance matrix using BP: condensing notation

$$\begin{aligned}
\kappa_j^i &\equiv J_{ij} v_j^{i\mathrm{BP}} - \frac{[m_j^{(i)\mathrm{BP}} - m_j^{i\ \mathrm{BP}}]}{m_i^{\mathrm{BP}}} && (30)\\
u_j^i &\equiv m_j^{(i)\mathrm{BP}} + m_i^{\mathrm{BP}} \kappa_j^i && (31)\\
v_i &\equiv \frac{s_i}{1 - s_i \left[\alpha_i^{\mathrm{BP}} + \left(m_i^{\mathrm{BP}}\right)^{-1} \sum_{l \in \partial i} J_{il} [m_l^{(i)\ \mathrm{BP}} - m_l^{i\ \mathrm{BP}}] \right]} && (32)\\
m_i &\equiv m_i^{\mathrm{BP}} && (33)
\end{aligned}$$

we have the following equations:

$$\begin{aligned}
\langle \sigma_i^2 \rangle &= v_i + m_i^2 && (34)\\
\langle \sigma_i \sigma_j \rangle &= m_i u_j^i + v_i \kappa_j^i && (35)\\
\langle \sigma_j \sigma_k \rangle &= u_j^i u_k^i + v_i \kappa_j^i \kappa_k^i + \langle \sigma_j \sigma_k \rangle^{(i)} && (36)
\end{aligned}$$

These equations suggest inverting matrices by calculating correlation matrices on growing graphs might be a useful application. By subsequently attaching new variables to the graph and running BP, one finds the full correlation matrix with N runs of BP, just as with the procedure described in [1], but the cost of the BP runs is halved since the graph is growing along with the BP runs. However, we should not overlook the fact that the equations above introduce large number of additions and multiplications, such that in the end the total computational complexity for inverting a sparse matrix is similar to other well-known methods.

6 Nonlinear models: connections with EP

The fact that loop corrections in the above form are able to correct for the total BP error in the (co)variances is of course due to the Gaussian nature of the model. In discrete models, exact parameterizations of the full distribution by use of local marginals only is in general not possible, but loop corrections are able to increase the accuracy of the Bethe approximation. Thus the above formalism might seem a promising basis for extensions to models that are not exactly tractable, possibly as an alternative for related algorithms like Expectation Propagation (EP) [9]. Since BP may be viewed as a special case of EP, we may hope for some generalizations of loop corrections equations, with some relation to EP, applicable in cases where function approximations become necessary. However, the specific form of EP equations very much depends on the choice of the approximating family of functions one chooses. The equivalence with BP corresponds to a family of approximate EP functions that fully factorizes over the variables of the model [9]. For loop corrected BP strategies, we expect a relationship with EP approaches based on larger local neighborhoods.

We will investigate the relation to EP by deriving equations for models with general nonlinear single-variable potentials, i.e., $\psi_i(\sigma_i) \to \psi_i(\sigma_i)e^{-V_i(\sigma_i)}$, as one might expect in vision problems with nonlinear observation functions.

6.1 Full Gaussian EP

The Gaussian loop corrections approach seems rather similar to an EP approach where one includes a full Gaussian in the approximating target distribution. The standard EP formalism for this approach is to choose as an approximate distribution

$$q(\boldsymbol{x}) \sim \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{m})\right] \tag{37}$$

where

$$\Sigma^{-1} = \Sigma_{\mathrm{g}}^{-1} + \sum_i (\Sigma^i)^{-1} \tag{38}$$

$$\Sigma^{-1}\boldsymbol{m} = \Sigma_{\mathrm{g}}^{-1}\boldsymbol{m}_{\mathrm{g}} + \sum_i (\Sigma^i)^{-1}\boldsymbol{m}^i \tag{39}$$

and the subscript g stands for Gaussian, as it represents the Gaussian contribution to the full joint probability. The remaining terms relate to approximations of the single node potentials in the following manner

$$q(\boldsymbol{x}) = q_g(\boldsymbol{x}) \prod_i \overline{f}^i(\boldsymbol{x}) \tag{40}$$

Here again $q_g(\boldsymbol{x}) = \exp[-(\boldsymbol{x}-\boldsymbol{m}_g)^T \Sigma_g^{-1}(\boldsymbol{x}-\boldsymbol{m}_g)/2]$ and $\overline{f}^i(\boldsymbol{x})$ is the standard Minka notation [9] for a term that approximates an intractable contribution, in our case

$$f^i(\boldsymbol{x}) = f^i(x_i) = e^{-V_i(x_i)} \tag{41}$$

Updating the parameters $\boldsymbol{m}^i$ and Σ^i proceeds in the usual way: first, for a term i, the contribution of its approximation is removed from the full joint:

$$q^{\setminus i}(\boldsymbol{x}) \sim \frac{q(\boldsymbol{x})}{\overline{f}^i(\boldsymbol{x})} \tag{42}$$

meaning that

$$(\Sigma^{\setminus i})^{-1} = \Sigma_{\mathrm{g}}^{-1} + \sum_{j(\neq i)} (\Sigma^j)^{-1} \tag{43}$$

$$(\Sigma^{\setminus i})^{-1}\boldsymbol{m}^{\setminus i} = \Sigma_{\mathrm{g}}^{-1}\boldsymbol{m}_{\mathrm{g}} + \sum_{j(\neq i)} (\Sigma^j)^{-1}\boldsymbol{m}^j \tag{44}$$

Then the new value of the full parameters is obtained by defining

$$\hat{p}(\boldsymbol{x}) = \frac{q^{\setminus i}(\boldsymbol{x}) f^i(x_i)}{\int d\boldsymbol{x} q^{\setminus i}(\boldsymbol{x}) f^i(x_i)} \tag{45}$$

and minimizing

$$\begin{aligned} KL(\hat{p}|q) &= \int d\boldsymbol{x}\hat{p}(\boldsymbol{x}) \log\left[\frac{\hat{p}(\boldsymbol{x})}{q(\boldsymbol{x})}\right] \\ &\sim \int d\boldsymbol{x} q^{\setminus i}(\boldsymbol{x}) f^i(x_i) \log\left[\frac{q^{\setminus i}(\boldsymbol{x}) f^i(x_i)}{q_1(\boldsymbol{x}) q_2(x_i)}\right] \\ &= \int d\boldsymbol{x} q^{\setminus i}(\boldsymbol{x}) f^i(x_i) \left\{\log\left[\frac{f^i(x_i)}{q_2(x_i)}\right] + \log\left[\frac{q^{\setminus i}(\boldsymbol{x})}{q_1(\boldsymbol{x})}\right]\right\} \end{aligned} \tag{46}$$

where we have taken the liberty of decomposing the Gaussian function $q(\boldsymbol{x})$ into a Gaussian that depends only on x_i and a remaining Gaussian depending on the whole vector $\boldsymbol{x}$. Since

both $q_1(\boldsymbol{x})$ and $q^{\backslash i}(\boldsymbol{x})$ are Gaussians, the KL-divergence is minimal when they are equal and thus we have to minimize

$$KL(\hat{p}|q) = \int dx_i q^{\backslash i}(x_i) f^i(x_i) \log \frac{f^i(x_i)}{q_2(x_i)} \tag{47}$$

with respect to $q_2(x_i)$, where $q^{\backslash i}(x_i) = \int d\boldsymbol{x}_{\backslash x_i} q^{\backslash i}(\boldsymbol{x})$ and it is clear that $q_2(x_i)$ is parameterized by $(\Sigma^i)^{-1}$ and m^i, the only parameters to be updated. We furthermore deduce that these parameters contribute only to single entries in the matrices and vectors (i.e. they are scalars). Thus

$$m^i = Z^{-1} \int dx_i \; x_i \; q^{\backslash i}(x_i) f^i(x_i) \tag{48}$$

$$\Sigma^i = Z^{-1} \int dx_i \; x_i^2 \; q^{\backslash i}(x_i) f^i(x_i) - (m^i)^2 \tag{49}$$

$$Z = \int dx_i \; q^{\backslash i}(x_i) f^i(x_i) \tag{50}$$

The marginalization of $q^{\backslash i}(\boldsymbol{x})$ yields,

$$q^{\backslash i}(x_i) \sim \exp\left[-\frac{(x_i - m_i^{\backslash i})^2}{2\Sigma_{ii}^{\backslash i}}\right] \tag{51}$$

where

$$\Sigma^{\backslash i} = \left[(\Sigma_{\mathrm{g}})^{-1} + \mathrm{diag}_{\backslash i}\left(\frac{1}{\Sigma^j}\right)\right]^{-1} \tag{52}$$

$$m_i^{\backslash i} = \sum_l (\Sigma^{\backslash i})_{il} \left[[(\Sigma_{\mathrm{g}}^{-1})\boldsymbol{m}_{\mathrm{g}}]_l + \frac{m^l}{\Sigma^l}(1 - \delta_{il})\right] \tag{53}$$

Thus the most costly computations are the inversion in equation (52) and the one-dimensional integrations of (50). For very large models the inversions may become prohibitive, but otherwise this scheme seems efficient, since the "cavity covariances" do not have to be computed separately but are implicitly present in these inversions, and are optimal with respect to the minimization of the KL-divergence.

6.2 Loop corrections formulation

In this subsection we will discuss a possible generalization of the loop correction scheme for the model discussed in the previous subsection. The same model with nonlinear single-variable potentials may be tackled starting from the loop correction scheme at the beginning of this

paper, by slightly generalizing it to the case where $\psi_i(\sigma_i) \to \psi_i(\sigma_i)e^{-V_i(\sigma_i)}$. The formalism at the beginning of this paper may still be applied when the Gaussian parameterization of the distributions $P^{(i)}(\boldsymbol{\sigma}_{\partial_i})$ for all i is an approximation. For given estimates of the covariance matrices A_i we then find:

$$
\begin{aligned}
m_i^j &= \frac{\int d\sigma\sigma \exp\left\{-\Phi_i^j(\sigma, \mathbf{m}^i, \alpha_i^j(A_i, \{v_l^i\}))\right\}}{\int d\sigma \exp\left\{-\Phi_i^j(\sigma, \mathbf{m}^i, \alpha_i^j(A_i, \{v_l^i\}))\right\}} \\
&\quad - \frac{\epsilon_j^i \int d\tau\tau \exp\left\{-\Phi_j^i(\tau, \mathbf{m}^j, \alpha_j^i(A_i, \{v_l^i\}))\right\}}{\int d\tau \exp\left\{-\Phi_j^i(\tau, \mathbf{m}^j, \alpha_j^i(A_i, \{v_l^i\}))\right\}} && (54) \\
v_i^j &= \frac{\int d\sigma\sigma^2 \exp\left\{-\Phi_i^j(\sigma, \mathbf{m}^i, \alpha_i^j(A_i, \{v_l^i\}))\right\}}{\int d\sigma \exp\left\{-\Phi_i^j(\sigma, \mathbf{m}^i, \alpha_i^j(A_i, \{v_l^i\}))\right\}} \\
&\quad - \frac{\int d\tau(m_i^j + \epsilon_j^i\tau)^2 \exp\left\{-\Phi_j^i(\tau, \mathbf{m}^j, \alpha_j^i(A_i, \{v_l^i\}))\right\}}{\int d\tau \exp\left\{-\Phi_j^i(\tau, \mathbf{m}^j, \alpha_j^i(A_i, \{v_l^i\}))\right\}} && (55) \\
m_i &= \frac{\int d\sigma\sigma \exp\left\{-\Phi_i(\sigma, \mathbf{m}^i, \alpha_i(A_i, \{v_l^i\}))\right\}}{\int d\sigma \exp\left\{-\Phi_i(\sigma, \mathbf{m}^i, \alpha_i(A_i, \{v_l^i\}))\right\}} && (56) \\
v_i + m_i^2 &= \frac{\int d\sigma\sigma^2 \exp\left\{-\Phi_i(\sigma, \mathbf{m}^i, \alpha_i(A_i, \{v_l^i\}))\right\}}{\int d\sigma \exp\left\{-\Phi_i(\sigma, \mathbf{m}^i, \alpha_i(A_i, \{v_l^i\}))\right\}} && (57)
\end{aligned}
$$

with

$$
\begin{aligned}
\Phi_i^j(\sigma, \mathbf{m}^i, \alpha_i^j(A_i, \{v_l^i\})) &= \frac{(\sigma - \hat{m}_i^j)^2}{2\hat{v}_i^j} + V_i(\sigma) && (58) \\
\hat{v}_i^j &= \frac{s_i}{1 - s_i\alpha_i^j(A_i, \{v_l^i\})} && (59) \\
\hat{m}_i^j &= \hat{v}_i^j \left[\frac{\mu_i}{s_i} + \sum_{k\in\partial i\setminus j} J_{ik} m_k^i\right] && (60) \\
\Phi_i(\sigma, \mathbf{m}^i, \alpha_i(A_i, \{v_l^i\})) &= \frac{(\sigma - \hat{m}_i)^2}{2\hat{v}_i} + V_i(\sigma) && (61) \\
\hat{v}_i &= \frac{s_i}{1 - s_i\alpha_i(A_i, \{v_l^i\})} && (62) \\
\hat{m}_i &= \hat{v}_i \left[\frac{\mu_i}{s_i} + \sum_{k\in\partial i} J_{ik} m_k^i\right] && (63)
\end{aligned}
$$

On the one hand, it is easy to check that these equations reduce to the BP equations with loop correction when $V_i(\sigma_i) = 0$ for all i, i.e. $m_i^j = \hat{m}_i^j - \epsilon_j^i \hat{m}_j^i$ and $v_i^j = \hat{v}_i^j - (\epsilon_j^i)^2 \hat{v}_j^i$. In that case they should be equivalent to the full Gaussian EP approach of the previous subsection as well, since both treatments are exact in this limit. On the other hand, when we take $V_i(\sigma_i) \neq 0$ and $A_i = 0$ for all i, these updates are somewhat similar to EP with completely factorizing Gaussian approximate target distribution (i.e., deriving equations starting from diagonal Σ). The slight difference is due to the fact that the propagated expectations m_i^j and v_i^j parameterize approximate cavity distributions (i.e. in absence of one neighboring variable) and not the actual target marginal distributions. Thus the KL-divergence with an approximate factorizing cavity distribution is minimized and not with the approximate target distribution. When $V_i(\sigma_i) = 0$ for all i, this difference vanishes, and the algorithm reduces to EP with a factorizing Gaussian as approximate joint distribution, which, in that limit (where integrals may be performed exactly) is equivalent to ordinary BP.

When optimizing the approximations for marginal moments is the objective of the algorithm, the approach of this subsection is obviously not optimal, since instead moments of cavity distributions are optimized in the integrals that calculate the messages.

6.3 Alternative loop correction formalism

Inspired by the above observations regarding the optimization of the marginal moments of the target approximation, one may derive alternative consistency equations as in [7], starting from the expressions for the actual marginals, such that the integrations include full sets of neighboring factors. Once again, we approximate the cavity distributions by Gaussians, and find

$$m_i^j = \langle \sigma_i \rangle_{\hat{i}} - [J_{ij} v_i^j + \epsilon_j^i] \langle \sigma_j \rangle_{\hat{j}} \tag{64}$$

$$v_i^j = \left[\langle \sigma_i^2 \rangle_{\hat{i}} - \left(\langle \sigma_i \rangle_{\hat{i}} \right)^2 \right] - \left(J_{ij} v_i^j + \epsilon_j^i \right)^2 \left[\langle \sigma_j^2 \rangle_{\hat{j}} - \left(\langle \sigma_j \rangle_{\hat{j}} \right)^2 \right] \tag{65}$$

with

$$\langle \sigma_i \rangle_{\hat{i}} = \frac{\int d\sigma_i \sigma_i \exp\left[-\Phi_i(\sigma_i, \mathbf{m}^i, \alpha_i(A_i, \{v_l^i\})) \right]}{\int d\sigma_i \exp\left[-\Phi_i(\sigma_i, \mathbf{m}^i, \alpha_i(A_i, \{v_l^i\})) \right]} \tag{66}$$

$$\langle \sigma_i^2 \rangle_{\hat{i}} = \frac{\int d\sigma_i \sigma_i^2 \exp\left[-\Phi_i(\sigma_i, \mathbf{m}^i, \alpha_i(A_i, \{v_l^i\})) \right]}{\int d\sigma_i \exp\left[-\Phi_i(\sigma_i, \mathbf{m}^i, \alpha_i(A_i, \{v_l^i\})) \right]} \tag{67}$$

For $A_i = 0$ this reduces to EP with fully factorizing Gaussian, and again $V_i(\sigma_i) = 0$ leads to equations which are equivalent to BP. A suitable choice of A_i should make the fixed points of the above equations equivalent to the full-Gaussian EP equations at the beginning of this section, since both approaches optimize the marginal moments of each variable, given a Gaussian

interaction matrix with the rest of the model. However, the benefit of full Gaussian EP is that this Gaussian interaction matrix is optimized on the way, albeit at the cost of an inversion at each iteration, while the loop corrected approach desires an estimate of A_i as input, which is not further updated.

Thus loop corrections are an alternative for the current type of model only if these inversions are so costly that approximations of the above form are sensible.

6.4 Estimating A_i: response propagation

In the above formalism, an approximation for the cavity covariance matrix A_i may be obtained by applying a linear response algorithm to the graph from which variable i has been removed. The entries of A_i for variables $j, k \in \partial_i$ follow from

$$\langle \sigma_j \sigma_k \rangle^{(i)} - \langle \sigma_j \rangle^{(i)} \langle \sigma_k \rangle^{(i)} = s_k \frac{\partial \langle \sigma_j \rangle^{(i)}}{\partial \mu_k} \tag{68}$$

$$= s_k \frac{\partial m_j^{(i)}}{\partial \mu_k} \tag{69}$$

Thus derivatives of average messages should be computed. Using (57) and (63), we estimate them by taking the derivatives of these update equations neglecting the corresponding cavity covariances (setting $A_j = 0$ for all $j \neq i$):

$$\frac{\partial m_j^{(i)}}{\partial \mu_k} = v_j^{(i)} \left[\frac{\delta_{jk}}{s_j} + \sum_{l \in \partial_j} J_{jl} \frac{\partial m_l^{j(i)}}{\partial \mu_k} \right] \tag{70}$$

$$\frac{\partial m_l^{j(i)}}{\partial \mu_k} = v_l^{j(i)} \left[\frac{\delta_{lk}}{s_l} + \sum_{n \in \partial_l \setminus j} J_{ln} \frac{\partial m_n^{l(i)}}{\partial \mu_k} \right] \tag{71}$$

The reasoning here is that an estimate of the covariance matrix in “zeroth order” enables a first order corrected version (see reference [5]) of the expectation propagation algorithm, equations (57) and (63) with $A_i = 0 \; \forall i$. Note that, given the values of the single node variances, which have to be obtained via running EP including integrations on the graph without variable i, we have a fast algorithm for the responses, that does not involve integrations. Thus the cost for a cavity covariance matrix is determined by the factorizing EP updates on the graph without the central variable i, after which the responses may quickly be obtained. This way of estimating A_i is obviously as costly as a number of inversions of matrices of dimension N In the full Gaussian EP approach, an inversion is necessary at each update, such that in the end the present approach might become cheaper, but for the current model EP itself seems more practicle. Possibly however, for more complex models the loop correction approach is beneficial, a topic which is to be further investigated.

7 Discussion

In this paper we have derived loop corrected belief propagation equations for continuous variable models. In particular we have worked out the exactly tractable case of Gaussian models, and have derived the exact message passing equations. The role of the "connected correlations" of the cavity distribution discussed in [5] is taken over by the off-diagonal parts of the covariance matrix, which may be obtained in preprocessing algorithms. Moreover, using the fact that if ordinary BP converges, it produces exact marginal averages, various relations between BP messages are obtained, leading to alternative update schemes to invert the covariance matrix.

For models involving nonlinear terms, for which approximation algorithms are needed in order to compute marginals, we discussed some relations between expectation propagation approaches and loop correction strategies, in particular for a model class where the nonlinear potentials involve only single variables. Loop correction approaches for continuous variable models become attractive once alternative strategies like expectation propagation grow too costly. On the other hand, however, they are themselves heavily penalized by the cost of the preprocessing stage necessary to estimate the cavity connected correlations.

References

[1] Welling M. and Teh Y.W.*Neural Comput* **16**(1) (2004) 197-221.

[2] Pellizola A., *J. Phys. A: Math. Gen.***38** (2005) R309

[3] Minka T., Qi Y. in *Advances in Neural Information Processing Systems* 2004

[4] Minka T., Welling M. and Teh Y.W. in *21st Conference on Uncertainty in Artificial Intelligence (UAI)*, (2005)

[5] Montanari A and Rizzo T., *J. Stat. Mech.*(2005) P10011

[6] Rizzo T., Wemmenhove B. and Kappen H.J. cond-mat/0608312 to appear in *Phys. Rev. E*

[7] Mooij J.M. and Kappen H.J. *Journal of Machine Learning Research* **8**(2007) 1113-1143,

[8] Weiss Y. and Freeman W.T. *Neural Computation***13** (2001) 2173-2200

[9] Minka T, *A family of algorithms for approximate Bayesian inference* (phd thesis).

A Full covariance matrix as a function of BP quantities

Directly writing out the error $\alpha_i^{\mathrm{LC}} - \alpha_i^{\mathrm{BP}}$ results in

$$\alpha_i^{\mathrm{LC}} - \alpha_i^{\mathrm{BP}} \quad = \quad \sum_{k,l \in \partial i} J_{il} \left[\delta_{kl} [v_l^{i\ \mathrm{LC}} - v_l^{i\ \mathrm{BP}}] + (A_i)_{kl}(1 - \delta_{kl}) \right] J_{ik} \tag{72}$$

It follows that

$$J_{il}[v_l^{i\ \mathrm{LC}} - v_l^{i\ \mathrm{BP}}] + \sum_{k \in \partial i \backslash l} J_{ik}(A_i)_{kl} = -\frac{m_l^{(i)\ \mathrm{BP}} - m_l^{i\ \mathrm{BP}}}{m_i^{\mathrm{BP}}} \tag{73}$$

from which we may furthermore derive the relation

$$\{(D_i + A_i)\mathbf{J}_i\}_j = J_{ij} v_j^{i\mathrm{BP}} - \frac{[m_j^{(i)\mathrm{BP}} - m_j^{i\ \mathrm{BP}}]}{m_i^{\mathrm{BP}}} \tag{74}$$

A.1 Off-diagonal parts of covariance matrix

The off-diagonal elements of the covariance matrix of the model may be expressed in terms of cavity distributions again, using the formulation of subsection (2.2). We find, using the measures (10),

$$\begin{aligned} \langle \sigma_i \sigma_j \rangle \quad &= \quad \mu_i \langle \sigma_j \rangle_i + s_i \sum_{l \in \partial_i} J_{il} \langle \sigma_l \sigma_j \rangle_i \\ &= \quad m_i^{\mathrm{BP}} m_j^{(i)\ \mathrm{BP}} + [(m_i^{\mathrm{BP}})^2 + v_i^{\mathrm{LC}}] \left[J_{ij} v_j^{i\ \mathrm{BP}} - \frac{[m_j^{(i)\mathrm{BP}} - m_j^{i\ \mathrm{BP}}]}{m_i^{\mathrm{BP}}} \right] \end{aligned} \tag{75}$$

where we have used (74).

Next nearest neighbor correlations follow from a similar calculation as above:

$$\begin{aligned} \langle \sigma_j \sigma_k \rangle \quad &= \quad \langle \sigma_j \sigma_k \rangle_i \qquad\qquad\qquad j, k \in \partial_i \\ &= \quad m_j^{(i)\ \mathrm{BP}} m_k^{(i)\ \mathrm{BP}} + m_i^{\mathrm{BP}} \left[m_j^{(i)\ \mathrm{BP}} \{(D_i + A_i)J_i\}_k + m_k^{(i)\ \mathrm{BP}} \{(D_i + A_i)J_i\}_j \right] \\ &\quad + [(m_i^{\mathrm{BP}})^2 + v_i^{\mathrm{LC}}] \{(D_i + A_i)J_i\}_j \{(D_i + A_i)J_i\}_k + (D_i + A_i)_{jk} \end{aligned} \tag{76}$$

All terms follow from BP on the graph with and without i, except for the last one. Using (74), and renaming terms, we obtain equations (36).